2024 年第 2 期（总第 37 期）

国际战略与安全研究报告

INTERNATIONAL SECURITY AND STRATEGY STUDIES REPORT

人工智能术语研究阶段性成果

清华大学战略与安全研究中心
CENTER FOR INTERNATIONAL SECURITY AND STRATEGY
TSINGHUA UNIVERSITY

人工智能术语研究阶段性成果

CISS人工智能与国际安全项目术语工作组

一、武器（系统）

1. 武器

泛指用于攻击或保卫自身的作战器械与装置，被用于直接杀伤敌人的有生力量、破坏敌方设施的器械、装置，或在遭受进攻时进行防御的器械、装置，可区分为攻击性武器与防御性武器两类。

武器装备：武器装备是作战人员用于执行和保障作战行动的武器、武器系统以及相关的其他军事技术装备的统称，是进行战争的重要物质基础，是军队战斗力生成的重要力量，是完成各种军事行动的重要支撑。[①]

军械：军械主要包括枪械、火炮、弹药、战术导弹、光学仪器（侦察器材，如望远镜、潜望镜、炮队镜、高炮指挥镜等）、瞄准器材、测角器材、通用雷达、指挥仪、移动电站、防暴器材、冷兵器[②]等一系列用于军事目的的武器与装备。

弹药：指含有火药、炸药或其他填充物，能够对目标造成损害或完成其他战术任务的物品。这包括炮弹、火炮炮弹、手榴弹、步枪榴弹、航空炸弹、火箭、导弹、鱼雷、水雷、地雷等，以及用于狩猎和射击运动等非军事用途的礼炮和弹药。[③]

---

① *The people's Liberation Army military China,*《中国人民解放军军语》，中国总参谋部、总政治部、总后勤部、总装备部发布，2011.

② 中国军事百科全书编审委员会. (2016).《中国军事百科全书》. 中国大百科全书出版社.

③ 中国军事百科全书编审委员会. (2016).《中国军事百科全书》. 中国大百科全书出版社.

装备：给军队配备的武器、军装、器材、技术力量等。④

2.武器系统：

由若干功能上相互关联的武器、技术装备等有序组合，协同完成一定作战任务的有机整体。⑤

不可接受的致命性自主武器系统：根据《中国关于“致命性自主武器系统”问题的工作文件》⑥，中国认为“不可接受的”自主武器系统应满足但不限于以下五个基本特征：

（1）致命性：即具有足以致命的载荷和手段；

（2）全自主性：即在执行任务整个过程中均无人的介入和控制：

（3）无法终止性：即启动后没有终止手段；

（4）滥杀性：即不分条件、场合和对象，自动执行杀伤任务；

（5）进化性：即在与环境交互过程中，通过自主学习，实现功能扩展和能力进化，且超出人的预测。

“可接受的”自主武器系统：具有较高的自主程度，但应始终处于人类控制之下，且可被安全、可信、可靠、可控地使用，人类可随时中止其运行。在军事行动中应能够遵循区分、比例、预防等国际人道主义法基本原则。⑦

3.非致命性武器

非致命性武器是军队等强力部门在执勤、处突、反恐等任务中广

④ 中国军事百科全书编审委员会. (2016).《中国军事百科全书》. 中国大百科全书出版社.

⑤ https://www.term.org.cn/CN/abstract/abstract11322.shtml

⑥ https://documents.unoda.org/wp-content/uploads/2022/07/Working-Paper-of-the-Peoples-Republic-of-China-on-Lethal-Autonomous-Weapons-SystemsChinese.pdf

⑦ 中国关于“致命性自主武器系统”问题的工作文件，2022年7月，https://documents.unoda.org/wp-content/uploads/2022/07/Working-Paper-of-the-Peoples-Republic-of-China-on-Lethal-Autonomous-Weapons-SystemsChinese.pdf

泛使用的武器，在不致命的前提下使目标快速失能并且失能后果具有最大概率的可逆性。非致命性武器提供了以最低限度的武力开展行动的能力，控制了暴力和防止不必要的附带伤亡和破坏。⑧

二、智能/自主平台（系统）

4.机器人

自动执行工作的机器装置。既可以接受人类指挥，又可以运行预先编排的程序，也可以根据以人工智能技术制定的原则纲领行动。机器人的主要任务是协助或取代人类的工作，例如生产业、建筑业，或是危险的工作。⑨

5.智能武器

是指配备了人工智能（AI）、传感器、自动化系统等使其具备一定程度自主决策、目标识别、跟踪和攻击能力，用于实施和保障战斗行动的武器、武器系统和与之配套的其他军事技术器材的统称。这些武器能够在复杂的战场环境中不同程度地提高作战效率和精度，并减少对人为操作的依赖。所谓智能是对人的能力延伸和发展，是人机交互环境中的一部分。人作为军事活动和智能活动的主体，对自主化程度较高的智能武器系统应拥有必要的干预能力。⑩

6.自主武器系统

关于“自主性”，其主要目的应是降低军事行动中武器系统对人类及外部资源的依赖程度，提高对复杂动态环境的适应性和战场生存

---

⑧ 赵陕东，马永忠.非致命武器与警用器材[M].北京：兵器工业出版社，2005：45-46.

⑨ 全国科学技术名词审定委员会：《计算机科学技术名词（第三版）》，科学出版社，2018年，第460页。

⑩ 结合过去20年中国学者观点，并基于智能武器基本逻辑进行编纂而成。

性，从而更好地完成人类赋予的任务。应根据不同场景、不同程度的自主能力，有针对性规范相关武器系统的使用。如果自主能力未用于杀伤链（例如用于情报搜集和侦察的无人机），即使一些武器系统自主程度很高，其所具有的自主性也不会引发人道主义关切。

武器系统杀伤链包括观察、判断、决策、行动等多个关键环节，在某些环节具有自主功能的武器系统并不必然导致滥杀滥伤。因此，笼统地推出禁止或限制措施或将损害各国正当国防能力，甚至损害各国和平利用相关技术的权利。各方应考虑将自主武器系统分为“不可接受的”和“可接受的”两类，对“不可接受的”部分加以禁止，对“可接受的”部分加以规范，以确保有关武器系统安全、可靠、可控，遵循国际人道法及其他适用的国际法。[11]

——《中国关于“致命性自主武器系统”问题的工作文件》

7.集群智能/群体智能

人工智能集群[12]：遵循统一控制的，人工智能计算功能单元的集合。

注1：人工智能计算功能单元可包含人工智能加速处理器、人工智能服务器、人工智能加速模组等。

注2：当由人工智能服务器组成时，人工智能集群可称为人工智能服务器集群，其中的人工智能服务器可称为节点。

群体智能[13]：群体智能（swarm intelligence, SI）也称为集群智能（collective intelligence, CI）……对于一个由众多简单个体组成的群体，

⑪ Working Paper of the People's Republic of China on Lethal Autonomous Weapons Systems, 9 August 2022, https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-_Group_of_Governmental_Experts_(2022)/CCW-GGE.1-2022-WP.6.pdf

⑫ 国家市场监督管理总局、国家标准化管理委员会：《中华人民共和国国家标准：信息技术 人工智能 术语（GB/T 41867-2022）》，2022年10月12日发布（2023年5月1日实施），第1页。

⑬ 张国辉、文笑雨：《群体智能》，清华大学出版社2022年，第1-3页。

若其个体具有能通过彼此间的简单合作来完成一个整体任务的能力，则称该群体具有“群体智能”……群体智能的核心是由众多简单个体组成的群体能够通过相互之间的简单合作来实现某一较复杂的功能，完成某一较复杂的任务。群体智能可以在没有 集中控制并且缺少全局信息和模型的前提下，为解决复杂的分布式问题提供了可能。

8.无人系统

无人系统指配备必要的数据处理单元、传感器、自动控制和通信系统，能够在无需搭载人类操作员的情况下发挥其力量执行指定任务的机电系统。

9.无人（作战）平台

无人作战平台是指在没有搭载人类操作人员的情况下执行作战任务的一种武器系统。

10.无人空中平台（无人机、无人驾驶航空器、遥控驾驶航空器）

无人驾驶航空器：是指没有机载驾驶员且自备动力系统的航空器。⑭

遥控飞机：由遥控站操作的无人驾驶飞机。遥控飞机是无人驾驶飞机的一个子类。⑮

11.无人地面平台（自动驾驶车辆、无人驾驶车辆）

无人地面平台是地面作战的重要力量，能够在没有搭载人类操作员的情况下执行各种任务。

---

⑭ 国务院、中央军委.《无人驾驶航空器飞行管理暂行条例》. 31 May 2023, www.gov.cn/zhengce/content/202306/content_6888799.htm.

⑮ Civil Unmanned Aircraft Systems Air Traffic Management Measures, https://jxj.beijing.gov.cn/zwgk/zcwj/bjszc/201911/t20191113_2674230.html

12.无人水面平台（无人水面艇）

无人水面舰艇是一种具备自主航行能力、通过搭载任务载荷来执行相关任务的水面舰艇。

13.无人潜航器（遥控水下潜航器、自主/自治水下潜航器）

无人水下航行器是一种无需搭载人类操作员，可以自动执行任务的水下航行器。

三、智能/自主技术

14.人工智能

人工智能是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。——《人工智能标准化白皮书（2018）》

人工智能已经进入全面反映人类智能的第三代技术。第一代利用知识、算法和算力三个要素构造，第二代利用数据、算法与算力三个要素构造。为了发展安全、可信、可靠与可扩展的AI技术，要同时利用知识、数据、算法和算力构造。——张钹、朱军等，《迈向第三代人工智能》

15.机器学习

机器学习指通过计算技术优化模型参数的过程，使模型的行为反映数据或经验。[16]作为一门学科，机器学习（Machine Learning）涉及统计学、系统辨识、逼近理论、神经网络、优化理论、计算机科学、脑科学等诸多领域。机器学习研究计算机怎样模拟或实现人类的学习

---

⑯ “Information technology - Artificial intelligence - Terminology” (GB/T 41867-2022) issued on May 1, 2023, Chinese Electronics Standardization Institute,https://std.samr.gov.cn/gb/search/gbDetailed?id=71F772D76866D3A7E05397BE0A0AB82A

行为，以获取新的知识或技能，重新组织已有的知识结构使之不断改善自身的性能，也是人工智能技术的核心。基于数据的机器学习是现代智能技术中的重要方法之一，研究从观测数据（样本）出发寻找规律，利用这些规律对未来数据或无法观测的数据进行预测。⑰

16. 深度学习

深度学习：通过训练具有许多隐层的神经网络来创建丰富层次表示的方法。

注：深度学习是机器学习的一个子集。——《国家标准丨GB/T 41867–2022》

深度学习：根据学习方法可以将机器学习分为传统机器学习和深度学习。深度学习是建立深层结构模型的学习方法，典型的深度学习算法包括深度置信网络、卷积神经网络、受限玻尔兹曼机和循环神经网络等。深度学习又称为深度神经网络（指层数超过3层的神经网络）。深度学习作为机器学习研究中的一个新兴领域，由Hinton等人于2006年提出。深度学习源于多层神经网络，其实质是给出了一种将特征表示和学习合二为一的方式。——《人工智能标准化白皮书（2018版）》

17. 神经网络

神经网络：由加权链路且权值可调整连接的基本处理元素的网络，通过把非线性函数作用到其输入值上使每个单元产生一个值，并把它传送给其他单元或把它表示成输出值。

注：虽然某些神经网络旨在模拟神经系统中神经元的功能，但大多数神经网络用于人工智能以实现连接模型。——《国家标准丨GB/T 5271.34–2006》

---

⑰ 中国电子技术标准化研究院 Chinese Electronics Standardization Institute. (2018). *White paper on artificial intelligence standardization*, 人工智能标准化白皮书（2018版）, http://www.cesi.cn/201801/3545.html

神经网络：由一层或多层神经元组成的网络，通过权值可调的加权连接，接收输入数据并产生输出。

注1：神经网络是连接主义方法的一个突出例子。

注2：虽然神经网络的设计最初是受生物神经元功能的启发，但大多数神经网络的研究不再遵循这种启发。——《国家标准丨GB/T 41867–2022》

18. 自主控制

自主控制是在没有人的干预下，把自主控制系统的感知能力、决策能力、协同能力和行动能力有机的结合起来，在非结构化环境下根据一定的控制策略自我决策并持续执行一系列控制功能完成预定目标的能力。

自动控制[18]：在没有人直接参与的情况下，利用外加的设备或装置，使机器、设备或生产过程的某个工作状态或参数自动地按照预定的规律运行。

自主控制系统[19]：通过计算机、伺服执行机构、传感反馈机构、稳定校正环节、数–模及模–数变换与分时采样通道、支持软件等构成的能实现自稳定、自调节、自录最优、自适应等功能的系统。

19. AI生命周期

AI生命周期指从AI项目概念阶段到部署和维护阶段结束的整个过程。中国在《人工智能标准化白皮书（2021版）》中引用国际标准化组织和国际电工组织第一联合技术委员会人工智能分委会 (ISO/IEC JTC 1 /SC 42)在ISO/IEC 22989《人工智能概念与术语》中提出了人工

⑱ 全国科学技术名词审定委员会:《计算机科学技术名词（第三版）》，科学出版社，2018年，第451页。

⑲ 卓名信、厉新光、徐继昌等主编:《军事大辞海（上）》，长城出版社，2000年，第1041页。

智能系统生命周期模型定义：

人工智能系统生命周期包括初始阶段、设计与开发、验证与确认、部署、运行与监测、重新评估及退出阶段。该生命周期模型源于系统和软件工程系统生命周期，并在此基础上强调了人工智能领域特性方面，包括开发运营、可追溯性、透明度及可解释性、安全与隐私、风险管理、治理等。[20]

四、行动与控制

20.指挥链

指挥链在中文语境中称指挥关系（command relationship）。通常情况下，指挥链或指挥关系是指从作战指挥高层延伸到基本作战单元的这样一条持续的职权线。根据等级、编制序列和任务的不同对指挥链上各级的指挥权限进行不同的职权划分，通常从上级到下级形成指挥与被指挥关系。根据作战需要或部队编制调整，可能会进行指挥关系转换，即改变原指挥关系，形成新指挥关系。

21.指挥控制（系统）

指挥控制是贯穿军事行动的一种重要工作。指挥员通过计划、指示、指挥控制系统等手段，掌握、调配、协调、制约作战部队和作战行动，以完成既定的作战任务。联合作战指挥控制，既包括对各种作战活动的掌控调控，也包括对各种行为主体职能行使和权责关系的规制约束。

指挥控制系统是保障指挥员和指挥机关对部队作战人员、作战行动、武器系统等实施指挥、控制、制约的系统。根据应用层级的不同，指挥控制系统可区分为战略、战役、战术级指挥控制系统。

---

[20]《人工智能标准化白皮书（2021版）》，中国电子技术标准化研究院，https://www.cesi.cn/202107/7796.html

22.决策点

中文语境中与决策点相近的概念是定下决心（decision-making），它指的是指挥员在接受上级任务、领会上级意图后，结合对作战态势的判断而作出的行动决定。

23.有目的的人类控制

这不是一个官方军事术语，是学者在论文中提出的概念。我们的理解是，自主武器系统或致命性AI赋能武器系统在目标核查、锁定目标、实施打击前等关键任务点，由人类进行最后确认，而不是完全依靠传感器、自主武器系统、或致命性AI武器系统的预测或判断。

五、人机关系

24.人机交互

人机交互是指人类用户与软硬件单元结合的计算机系统间进行信息输入和输出的交互活动，以协作实现符合人类需求的任务目标。在人工智能技术的背景下，人机交互旨在研究和设计人类与计算机系统之间的互动方式，通过智能技术提升人—机—环境的交互效率，使人类与计算机系统之间的沟通更加自然和真实，其主要涉及的领域包括自然语言处理、语音识别、手势识别、情感计算、虚拟现实和增强现实等。[㉑]

25.可信赖的人工智能

发展人工智能应坚持“以人为本”理念，以增进人类共同福祉

㉑ 参考资料：《以人为中心的交互系统设计过程》，中华人民共和国国家标准GB/T 18976-2003/ ISO 13407: 1999；范向民、范俊君、田丰、戴国忠：《人机交互与人工智能：从交替浮沉到协同共进》，载《中国科学：信息科学》2019年第3期；中国人工智能学会组编：《人工智能导论》，中国科学技术出版社2018年版；中国人工智能学会：《中国人工智能系列白皮书——大模型技术（2023版）》，2023年9月。

为目标，以保障社会安全、尊重人类权益为前提，确保“智能向善”，保障个人隐私和数据安全，确保人工智能始终朝着有利于人类文明进步的方向发展。为此目的，不断提升人工智能系统的透明性、稳定性、可解释性、可预测性、公平性和可靠性，提升数据真实性和准确性，确保人工智能技术可审核、可监督、可追溯，确保人工智能始终处于人类控制之下。[22]

透明性：人工智能系统的决策机制、运作过程、数据使用、行为结果等能够被人类清晰地理解、解释并信任，主要可体现为算法的透明性、数据的透明性、决策的透明性等，旨在实现人工智能技术可解释、可预测、可追溯、可审查、可问责。透明性有助于所有利益攸关方清楚地了解人工智能系统的工作方式，但不得危害国家利益和人工智能系统利益攸关方的利益。[23]

可解释性：人工智能系统能够以清晰、易懂和有意义的方式向人类揭示其推理、决策和预测过程，使人类能够理解并信任其作出的行为或得出的结论。可解释性侧重于为人工智能系统作出的决策提供可理解的原因，而不是尝试为“实现必要的优越特性”提供理由。[24]

㉒ 参考资料：中央网信办：《全球人工智能治理倡议》，2023年10月；中国人工智能学会：《中国人工智能系列白皮书——大模型技术（2023 版）》，2023年9月；国家新一代人工智能治理专业委员会：《新一代人工智能治理原则——发展负责任的人工智能》，2019年6月。

㉓ 参考资料：外交部：《中国关于加强人工智能伦理治理的立场文件》，2022年11月；中国人工智能学会：《中国人工智能系列白皮书——大模型技术（2023 版）》，2023年9月；国家新一代人工智能治理专业委员会：《新一代人工智能治理原则——发展负责任的人工智能》，2019年6月；中国信息通信研究院：《人工智能行业自律公约（征求意见稿）》，2019年6月。

㉔ 参考文献：《信息技术 人工智能 术语》，中华人民共和国国家标准GB/T 41867-2022，2022 年 10 月；ISO/IEC 22989: 2022 (en) Information technology - Artificial intelligence - Artificial intelligence concepts and terminology；中国信息通信研究院：《人工智能行业自律公约（征求意见稿）》，2019年6月；中国信息通信研究院：《可信人工智能白皮书》，2021年7月。

可追溯性：人工智能系统在整个生命周期中各个环节的决策数据集、处理过程和输出结果能够被记录、追踪和回溯，使人工智能系统的决策结果可被人类理解和追踪。[25]

可靠性：人工智能系统在各种条件下持续有效地与其预设目标一致地运行并获得期望结果的能力，既要有一定程度抵抗恶意攻击的能力，也要有遇到严重问题时的退回机制。[26]

可预测性：人类能够预见并理解人工智能系统在不同条件下会产生何种输出和行为，使人类能够对于拟产生的输出作出可靠的假设，这要求人工智能系统的输出结果准确、可靠且可被重复，且能有效抵御漏洞和恶意攻击。[27]

可控性：又称为“人类控制”，是指人类的判断应融入人工智能系统的研发、部署和使用流程，使人类能够有效地指导、监督、评估、干预、中止或终止人工智能系统的行为，确保人工智能系统符合伦理、法律和安全标准，同时建立问责机制，确保人类是最终责任主体。[28]

---

[25] 参考文献：国家人工智能标准化总体组、全国信标委人工智能分委会：《人工智能伦理治理标准化指南（2023 版）》，2023年3月；《致命性自主武器系统领域的新兴技术问题政府专家组2019年会议报告》，2019年9月。

[26] 参考文献：国家人工智能标准化总体组、全国信标委人工智能分委会：《人工智能伦理治理标准化指南（2023 版）》，2023年3月；《信息技术 人工智能 术语》，中华人民共和国国家标准GB/T 41867-2022，2022年10月。

[27] 参考文献：国家人工智能标准化总体组、全国信标委人工智能分委会：《人工智能伦理治理标准化指南（2023 版）》，2023年3月；《信息技术人工智能术语》，中华人民共和国国家标准GB/T 41867-2022，2022年10月。

[28] 参考文献：外交部：《中国关于规范人工智能军事应用的立场文件》，2021年12月；外交部：《中国关于加强人工智能伦理治理的立场文件》，2022年11月；国家人工智能标准化总体组、全国信标委人工智能分委会：《人工智能伦理治理标准化指南（2023 版）》，2023年3月。

26.决策链

目标识别：根据各种传感器提供的目标信息，由人工智能系统判定目标所处的环境、目标类别和类型以及目标敌我属性等。[29]

目标确认：在识别目标后，人工智能系统对目标的身份和性质进行验证和确认，以确保识别结果的准确性和可靠性，并依据战争法和交战规则确认目标是合法的攻击对象。

行动授权：基于明确的规则、目的和限制条件，将特定的行动权限授予人工智能系统，使其能够在一定的范围内在人类的监督和控制下作出决策、采取行动或执行任务，确保任务安全且有效地完成。

决策确认：在人工智能系统作出决策后，通过一系列过程和机制对该决策进行再次审查、核实和验证，确认决策的合法性、合理性、可行性和潜在风险，保障决策能够实现预期效果，符合人类的利益和目标。

任务终止：结束或停止人工智能系统正在进行的任务。任务终止可能由多种原因导致，包括任务目标已实现、资源不足无法继续、出现不可抗力的情形导致任务无法推进、任务本身失去意义和价值等。

任务重置：将一项任务恢复到初始状态或者重新设定任务的条件、目标、步骤和资源，以新的方式或基于新的情况重新开始执行该任务。

发表日期：2024年8月30日

㉙ 熊武一、周家法主编：《军事大辞海》，长城出版社2000年版。

CISS人工智能与国际安全项目术语工作组成员名单：

肖　茜　清华大学战略与安全研究中心副主任、
　　人工智能国际治理研究院副院长
陈　琪　清华大学战略与安全研究中心副主任、国际关系学系教授
朱启超　国防科技大学国防科技战略研究智库主任、教授
谢海斌　国防科技大学智能科学学院教授
徐纬地　国防大学原战略研究所研究员
董　汀　清华大学战略与安全研究中心助理研究员
孙成昊　清华大学战略与安全研究中心助理研究员
李　强　中国政法大学军事法研究所所长、副教授
鲁传颖　清华大学战略与安全研究中心特约专家，
　　同济大学政治与国际关系学院教授
祁昊天　北京大学国际安全与和平研究中心副主任、助理教授
张　伶　原国防大学国家安全学院副教授
郑乐锋　清华大学战略与安全研究中心人工智能与国际安全项目专员
张　丁　清华大学战略与安全研究中心研究助理
刘　源　清华大学战略与安全研究中心研究助理

审编：肖茜　　签发：达巍

扫码关注我们

清华大学战略与安全研究中心编印

办公地点：北京市海淀区清华大学明理楼428房间

联系电话：010-62771388

http://ciss.tsinghua.edu.cn　邮箱：ciss@tsinghua.edu.cn